

# Optimal linear drift for the speed of convergence of an hypoelliptic diffusion

Arnaud Guillin, Pierre Monmarché

December 23, 2016

## Abstract

Among all generalized Ornstein-Uhlenbeck processes which sample the same invariant measure and for which the same amount of randomness (a  $N$ -dimensional Brownian motion) is injected in the system, we prove that the asymptotic rate of convergence is maximized by a non-reversible hypoelliptic one<sup>1</sup>.

## 1 Introduction

For a potential  $V : \mathbb{R}^N \rightarrow \mathbb{R}$  such that  $\int e^{-V} < \infty$ , consider  $\mu$  the associated Gibbs law, namely the probability measure with a density with respect to the Lebesgue measure proportional to  $e^{-V}$ . In order to compute expectations with respect to  $\mu$ , Markov Chain Monte Carlo (MCMC) algorithms are widely spread. Such an algorithm is based on an ergodic Markov process  $(X_t)_{t \geq 0}$  whose unique invariant law is  $\mu$ , so that for  $T$  large enough  $X_T$  is not far to be distributed according to  $\mu$ . The efficiency of the algorithm is directly linked to the rate of convergence of  $X$  toward its equilibrium, which is why a fair amount of work has been devoted to accelerating this convergence (see [11] and references within). In particular, since there are many possible Markov processes to sample the same equilibrium  $\mu$ , the question arises to choose the fastest, if any.

Along with those obtained from a Metropolis-Hasting procedure (see e.g. [15] and references within), one of the most classical Gibbs sampler is the Fokker-Planck diffusion that solves the SDE

$$dX_t = -\nabla V(X_t)dt + \sqrt{2}dB_t \quad (1)$$

where  $B$  is a standard  $N$ -dimensional Brownian motion. Its generator is

$$L = -\nabla V \cdot \nabla + \Delta$$

where we recall the generator of a Markov process  $X$  is formally defined by

$$Lf(x) = (\partial_t)_{|t=0} \mathbb{E}(f(X_t) \mid X_0 = x).$$

The Fokker-Planck diffusion is a reversible process in the sense its generator is self-adjoint in  $L^2(\mu)$ . This property is of theoretical interest but, from a practical point of view, reversible processes are usually not optimal with regard to their speed of convergence. A particular

---

<sup>1</sup>This paper has been published in *Electron. Commun. Probab. Vol. 21 (2016)*, followed by an erratum. The present version already contains the corrections indicated in the erratum.

technique for improving the convergence of  $X$  to  $\mu$  is to add a divergence-free (with respect to  $\mu$ ) drift  $b$ , namely to consider the SDE

$$dX_t = (-\nabla V(X_t) + b(X_t)) dt + \sqrt{2} dB_t \quad (2)$$

with  $b$  such that  $\nabla \cdot (be^{-V}) = 0$  where  $\nabla \cdot$  stands for the divergence operator. That way, the equilibrium is not affected, but the process is no longer reversible and the convergence is improved (cf. [8, 9, 1, 10]). It can be easily seen if we consider for example convergence in  $L^2(\mu)$ , where the spectral gap of the reversible dynamic is a lower bound for the speed of convergence of the non reversible one (just by comparison of Dirichlet forms), see [9].

Another way to improve the convergence is to consider a kinetic process  $(X, Y)$  where  $X$  is the position and  $Y = dX/dt$  is the velocity, which acts as an instantaneous memory (see [6, 15, 5]). For instance the Langevin diffusion

$$\begin{cases} dX_t = Y_t dt \\ dY_t = -\nabla V(X_t) dt - Y_t dt + \sqrt{2} dB_t \end{cases} \quad (3)$$

admits  $e^{-H}$  as an invariant measure where the Hamiltonian is  $H(x, y) = V(x) + \frac{1}{2}|y|^2$ . In particular, the first marginal of this equilibrium is  $\mu$ . The Langevin diffusion is non-reversible and moreover it is hypoelliptic. It has been observed in [15] that it may converge faster than the reversible Fokker-Planck diffusion in some applied problems. It recently regained much interest under the name Hamiltonian Monte Carlo methods [7].

For both dynamics (2) and (3) it is difficult for a general potential  $V$  to obtain sharp theoretical bounds on the rates of convergence (see [13] for consideration on this matter in the metastable case, namely the regime  $\varepsilon \rightarrow 0$  with the potential  $V_\varepsilon = \frac{1}{\varepsilon}V$  where  $V$  has several local minima). A particular simple situation is the case where  $V$  is quadratic or, in other words,  $\mu$  is a Gaussian measure. Of course MCMC algorithms are not really relevant in practice regarding sampling according to Gaussian measures, but then the exact rates of convergence for (2) and (3) are trackable (see [2, 12] and below).

In this context, the purpose of the present work is to answer the problem raised in [10], namely: for a given Gaussian law  $\mu$ , is it possible to find the optimal divergence-free linear drift one can add to (3) in order to obtain the largest rate of convergence? More generally, what is the largest rate of convergence one can get when sampling according to  $\mu$  using a (possibly hypoelliptic) Markov diffusion with linear drift and constant diffusion coefficients?

Obviously this question is ill-posed since the invariant measure of  $(W_t)_{t \geq 0} = (X_{\lambda t})_{t \geq 0}$  is still  $\mu$  for any  $\lambda > 0$ , and  $W$  goes  $\lambda$  times faster than  $X$  to equilibrium. Following [6] we will thus work under the additional assumption that the total amount of randomness instantaneously injected in the system (that is, the trace of the diffusion matrix) is prescribed.

In the following, we will first introduce the main notations and recall basic facts about generalized Ornstein-Uhlenbeck processes. In Section 2, we present our main results, giving a positive and definite answer to the problem raised in [10]. Section 3 is dedicated to the proofs of our main results, whereas Section 4 presents numerical illustration of our results and present some thoughts on the general case we wish to tackle in the future.

## Notations

In this whole work,  $\mathcal{M}_N(\mathbb{R})$  is the set of  $N \times N$  real matrices,  $\mathcal{S}_N^{\geq 0}(\mathbb{R})$  (resp.  $\mathcal{S}_N^{> 0}(\mathbb{R})$ ) the set of positive definite (resp. semi-definite) symmetric ones and  $\mathcal{A}_N(\mathbb{R})$  is the set of anti-symmetric ones. The spectrum of a matrix  $A$  is  $\sigma(A)$ , its trace is  $Tr(A)$ , its transpose is  $A^T$  and vectors are considered as column matrices, so that the scalar product  $x \cdot y$  is  $x^T y$ . Finally  $\Re(\lambda)$  stands for the real part of  $\lambda \in \mathbb{C}$  and  $diag(a_1, \dots, a_N)$  stands for the diagonal matrix with coefficients  $a_i$ .

## Basic facts about Ornstein-Uhlenbeck processes

We recall here some facts whose proofs and details may be found for instance in [2]. A generalized Ornstein-Uhlenbeck process (OUP) is any diffusion with a linear drift and a constant matrix diffusion. In other words in dimension  $N$  it is the solution of an SDE of the form

$$dX_t = AX_t dt + \sum_{j=0}^N \sigma_j dB_t^j$$

where  $A$  is a constant matrix, the  $\sigma_j$ 's are constant vectors and the  $B^j$ 's are 1-dimensional independent Brownian motions. A Markov process is an OUP if and only if its generator is of the form

$$L_{A,D}f(x) := (Ax)^T \nabla f(x) + \nabla \cdot (D \nabla f)(x) \quad (4)$$

where  $D = \frac{1}{2} \sum \sigma_j \sigma_j^T$  is a positive semi-definite matrix and  $\nabla \cdot$  stands for the divergence operator. Recall that a measure  $\mu$  is said invariant for  $X$  (or equivalently for  $L_{A,D}$ ) if  $\text{Law}(X_0) = \mu$  implies  $\text{Law}(X_t) = \mu$  for all  $t \geq 0$ . For an OUP, an invariant measure is necessarily a (possibly degenerated) Gaussian distribution.

On the other hand the process is hypoelliptic if and only if  $\text{Ker } D$  does not contain any non-trivial subspace which is invariant by  $A^T$ , and in that case an invariant measure is necessarily unique and non-degenerated (it has a positive density with respect to the Lebesgue measure on  $\mathbb{R}^N$ ). In the case where the invariant measure exists and is unique, its density  $\psi_\infty$  is the unique solution of  $L'_{A,D}\psi = 0$  where

$$L'_{A,D}f(x) = -(Ax)^T \nabla f(x) - \text{Tr}(A)f(x) + \nabla \cdot (D \nabla f)(x)$$

is the dual in the Lebesgue sense of  $L_{A,D}$ .

We will focus mainly on generalized Ornstein-Uhlenbeck processes which have a non degenerate Gaussian distribution as invariant probability measure. Let  $N \geq 1$ ,  $S \in \mathcal{S}_N^{>0}(\mathbb{R})$  and

$$\psi_\infty(x) = \frac{(\det S)^{\frac{1}{2}}}{(2\pi)^{\frac{N}{2}}} \exp\left(\frac{-x^T S x}{2}\right)$$

be the density of the (non-degenerated) Gaussian distribution with covariance matrix  $S^{-1}$ . The dual of  $L_{A,D}$  in  $L^2(\psi_\infty)$  is then

$$\begin{aligned} L_{A,D}^* f(x) &= \frac{1}{\psi_\infty(x)} L'_{A,D}(f \psi_\infty)(x) \\ &= -((2DS + A)x)^T \nabla f(x) + \nabla \cdot (D \nabla f)(x) \\ &= L_{-(2DS+A),D} f(x). \end{aligned}$$

Starting from an initial distribution  $\psi_0$ , the law  $\psi_t$  and the density with respect to equilibrium  $\frac{\psi_t}{\psi_\infty}$  at time  $t$  of an OUP generated by  $L_{A,D}$  are (weak) solutions of

$$\partial_t \psi_t = L'_{A,D} \psi_t \quad \text{and} \quad \partial_t \left( \frac{\psi_t}{\psi_\infty} \right) = L_{A,D}^* \left( \frac{\psi_t}{\psi_\infty} \right).$$

## 2 Main results

Let

$$\mathcal{I}(S) = \{(A, D) \in \mathcal{M}_{N \times N}(\mathbb{R}) \times \mathcal{S}_N^{\geq 0}(\mathbb{R}), \text{Tr } D \leq N, L'_{A,D} \psi_\infty = 0\}$$

be the set of drift/diffusions matrices such that  $\psi_\infty$  is invariant for the corresponding OUP and with at most the same amount of randomness injected in the system as the reversible dynamics with generator

$$L_{-S, I_N} f(x) = -(Sx)^T \nabla f(x) + \Delta f(x).$$

For  $A \in \mathcal{M}_{d \times d}(\mathbb{R})$  we write

$$\rho(A) = \inf\{-\Re(\lambda), \lambda \in \sigma(A)\}.$$

As we will see later, if  $(A, D) \in \mathcal{I}(S)$  with  $\rho(A) > 0$  then  $\psi_t \rightarrow \psi_\infty$  as  $t \rightarrow \infty$  for all  $\psi_0$ , which implies  $\psi_\infty$  to be the unique invariant measure of  $L_{A, D}$ , which is by consequence necessarily hypoelliptic.

Our main result identifies the maximum of  $\rho(A)$  under the constraint  $(A, D) \in \mathcal{I}(S)$ .

**Theorem 1.** *For all  $S \in \mathcal{S}_N^{>0}(\mathbb{R})$ ,*

$$\max\{\rho(A), (A, D) \in \mathcal{I}(S)\} = \max \sigma(S).$$

For an OUP with drift matrix  $A$  the rate of convergence to equilibrium is  $\rho(A)$  (see [12] and below). Hence Theorem 1 states that it is possible to sample an OUP that converges at rate  $\max \sigma(S)$  to  $\psi_\infty$  using the same amount of randomness as the classical reversible dynamics with generator  $L_{-S, I_N}$ , while the latter converges at rate

$$\rho(-S) = \min \sigma(S).$$

This should be compared to the results of Hwang, Hwang-Ma and Sheu [8] or of Lelièvre, Nier and Pavliotis [10] (the first one being anterior, and the second one giving more explicit bounds for the convergence) that reads

$$\max\{\rho(A), A \text{ s.t. } (A, I_N) \in \mathcal{I}(S)\} = \frac{\text{Tr} S}{N}$$

which is the arithmetic mean of all eigenvalues of  $S$ . On the other hand, for  $(A, D) \in \mathcal{I}(S)$ , the process is reversible if and only if  $A = -(2DS + A)$ , namely  $A = -DS$ , and for  $D = \frac{N}{\text{Tr} S - 1} S^{-1}$  this gives

$$\max\{\rho(A), (A, D) \in \mathcal{I}(S), L_{A, D}^* = L_{A, D}\} = \frac{N}{\text{Tr} S - 1},$$

which is the harmonic mean of the eigenvalues of  $S$ . Finally, considering  $(1 - \varepsilon)L_{A, D} + \varepsilon L_{-S, I_N}$  for any arbitrary  $\varepsilon$ ,

$$\sup\{\rho(A), (A, D) \in \mathcal{I}(S), D \text{ invertible}\} = \max \sigma(S).$$

Note that

$$\min \sigma(S) \leq \frac{N}{\text{Tr} S - 1} \leq \frac{\text{Tr} S}{N} \leq \max \sigma(S)$$

and that the equalities hold only when  $S$  is a homogeneous dilation, in which case no non-reversible dynamics can yield any improvement of the rate of convergence to equilibrium. On the other hand when the eigenvalues have different orders of magnitude (which means the problem is multi-scale; cf. [10, Fig. 4] where  $S$  has uniformly distributed coefficients in  $[0, 1]$ ), the improvement is already clear from the reversible diffusion (with  $D = I_N$ ) to the non-reversible (but still elliptic) ones, but yet it seems even more drastic when hypoelliptic dynamics are allowed. Obviously, the cost to pay for an optimal asymptotic speed of convergence is an initial delay for small times. Moreover, due to possibly large coefficients in the drift, a theoretically optimal continuous-time diffusion may lead, after discretization, to a not-so-efficient algorithm

implemented in practice. We will leave aside this consideration, and only concentrate on the continuous-time problem.

We will focus here on the convergence in the entropy sense, ensuring for example also convergence in total variation via Pinsker's inequality. However, the same line of reasoning will also work for  $L^2$  convergence or  $\Phi$ -entropies (see [4]). More precisely, for a measure  $\mu$ , denote by

$$\text{Ent}_\mu(h) := \int h \ln h d\mu - \left( \int h d\mu \right) \ln \left( \int h d\mu \right)$$

the entropy of a positive function  $h$  with respect to  $\mu$ . For the reversible elliptic OUP with generator  $L_{-S, I_N}$ , it is well known that for all  $h > 0$

$$\text{Ent}_{\psi_\infty} \left( e^{tL_{-S, I_N}^*} h \right) \leq e^{-2\rho(-S)t} \text{Ent}_{\psi_\infty}(h).$$

This is nothing else than an equivalent formulation of the Gaussian logarithmic Sobolev inequality of Nelson, see for example [3] and references therein.

For a general OUP, if  $(A, D) \in \mathcal{I}(S)$ , according to [12, Corollary 12] there exists a constant  $c \geq 1$  such that for all  $h > 0$

$$\text{Ent}_{\psi_\infty} \left( e^{tL_{A, D}^*} h \right) \leq c e^{-2\rho(A)t} \text{Ent}_{\psi_\infty}(h)$$

at least if  $A$  is diagonalizable (when it is not the case, a polynomial in  $t$  prefactor should be added). For a non-reversible yet elliptic OUP,  $c$  may be strictly greater than 1 due to a change of norm, exactly as when we consider the transport semi-group  $e^{tL_{A, 0}} f = f(e^{tA} \cdot)$  alone and write

$$|e^{tA} x|^2 \leq |Q|^2 |Q^{-1} e^{tA} Q|^2 |Q^{-1} x|^2 \leq e^{-2\rho(A)t} |Q|^2 |Q^{-1}|^2 |x|^2.$$

When the process is both non-reversible and non-elliptic, there are two reasons for  $c$  to be greater than 1: the change of norm for  $e^{tA}$ , and the initial regularization which is really slower than in the elliptic case. Indeed, the part of  $c$  which is due to slow regularization may badly behave with  $N$ . More precisely, since the optimal  $(A, D) \in \mathcal{I}(S)$  we will consider will be very degenerated (the rank of  $D$  being 1), [12, Remark p.16] yields a constant  $c$  of order  $N^{40N^2}$  which is, at the very least, absolutely awful. Of course this estimate is the result of a succession of rough bounds and a more careful (and involved) analysis could certainly refine it, but it is unclear whether the optimal bound is less than exponential with respect to  $N$ .

Fortunately this problem disappears if we start the dynamics with an elliptic one and then switch to the hypoelliptic optimal one, ensuring thus first a quick regularization property.

**Theorem 2.** *For any  $C > 1$  we can construct  $(A, D) \in \mathcal{I}(S)$  such that for all  $h > 0$ , with finite entropy, and for all  $t, t_0 > 0$  with  $t \geq t_0$ ,*

$$\text{Ent}_{\psi_\infty} \left( e^{(t-t_0)L_{A, D}^*} e^{t_0 L_{-S, I_N}} h \right) \leq C \frac{1}{2t_0 \min \sigma(S)} e^{-2(\max \sigma(S))(t-t_0)} \text{Ent}_{\psi_\infty}(h).$$

*Moreover it is possible to construct  $(A, D) \in \mathcal{I}(S)$  with  $\|A\|_F \leq 4N^2 \sqrt{\frac{(\max \sigma(S))^3}{\min \sigma(S)}}$  (where  $\|A\|_F = \sqrt{\text{Tr}(A^T A)}$  is the Frobenius norm) such that for all  $h > 0$ , with finite entropy, and for all  $t \geq t_0 > 0$*

$$\text{Ent}_{\psi_\infty} \left( e^{(t-t_0)L_{A, D}^*} e^{t_0 L_{-S, I_N}} h \right) \leq \frac{1}{t_0 \min \sigma(S)} e^{-2(\max \sigma(S))(t-t_0)} \text{Ent}_{\psi_\infty}(h). \quad (5)$$

It is thus possible to completely quantify the initial loss, which, due to the role of the initial warm up via the reversible diffusion, boils down to the change of norm in the energy.

**Remarks:**

- As a function of  $t_0$ , the rhs of (5) is minimal for  $t_0^{-1} = 2 \max \sigma(S)$ , for which

$$Ent_{\psi_\infty} \left( e^{(t-t_0)L_{A,D}^*} e^{t_0 L_{-S, I_N}} h \right) \leq 2e^{\frac{\max \sigma(S)}{\min \sigma(S)}} e^{-2t \max \sigma(S)} Ent_{\psi_\infty}(h).$$

- The proof furnishes an explicit, algorithmic construction of an optimal  $(A, D)$ , with  $D$  of rank 1. More generally, it is not hard to see from the proof that  $D$  can be chosen with a rank at most the dimension of the eigenspace associated to  $\max \sigma(S)$ .
- In order to sample according to an  $N$ -dimensional Gaussian measure  $\psi_\infty$ , we can always add an  $(N+1)^{th}$  auxiliary variable and sample, with a given amount of randomness and at a rate arbitrarily large, according to an  $N+1$ -dimensional Gaussian measure whose first  $N$ -dimensional marginal is  $\psi_\infty$ . Of course, this should be counteracted in practice by numerical problems due to the discretization of the dynamics.
- In the same line, we can also consider the question of the kinetic process (3). More precisely, given  $S \in \mathcal{S}_N^{>0}(\mathbb{R})$ , then the equilibrium of the  $2N$ -dimensional process  $(X, Y)$  that solves

$$\begin{cases} dX_t &= Y_t dt \\ dY_t &= -\nu S X_t dt - \frac{1}{\nu} Y_t dt + \sqrt{2} dB_t \end{cases} \quad (6)$$

is  $\psi_\infty(dx) \otimes \gamma_\nu(dy)$ , where  $\gamma_\nu(dy) = \frac{1}{\sqrt{2\pi\nu^2}} e^{-\frac{1}{2\nu^2}|y|^2} dy$  is the Gaussian distribution with covariance matrix  $\nu^2 I_N$ . If the aim is to sample according to  $\psi_\infty$ , then the choice for the second  $N$ -dimensional marginal is open. Hence, we have to chose  $\nu$  in order to maximize the rate of convergence to equilibrium of the whole process.

If  $v$  is an eigenvector of  $S$  associated to an eigenvalue  $\lambda$ , then  $(v, -rv)$  is an eigenvector of

$$A = \begin{pmatrix} 0 & I_N \\ -\nu S & -\frac{1}{\nu} I_N \end{pmatrix}$$

associated to the eigenvalue  $-r$  if and only if  $r^2 - \frac{1}{\nu} r + \lambda \nu = 0$ , namely

$$r = \frac{1}{2\nu} \left( 1 \pm \sqrt{1 - 4\lambda\nu^3} \right).$$

When  $4\nu^3 > \max \sigma(S^{-1})$ , the rate of convergence of  $(X, Y)$  is thus  $(2\nu)^{-1}$ , which goes to zero as  $\nu$  goes to  $\infty$ . On the other hand, when  $\nu$  goes to 0, this rate is equivalent to  $\nu^2 \min \sigma(S)$ , which again goes to 0. This proves that it is not possible to reach an arbitrarily large rate of convergence with the dynamics (6), and that the best choice for  $\nu$ , the variance of the velocity, is neither to be found at infinity nor at zero (which answers a question raised in [13]). Of course this would be a completely different story if we were to consider

$$\begin{cases} dX_t &= \frac{1}{\nu} Y_t dt \\ dY_t &= -S X_t dt - \frac{1}{\nu} Y_t dt + \sqrt{2} dB_t, \end{cases}$$

which also allows to sample according to  $\psi_\infty$ , but is no longer a kinetic process in the sense  $Y$  is no more the velocity  $\frac{dX}{dt}$ .

The optimal value of  $\nu$  in (6) is explicit if  $S = \lambda I_N$  is a homothety. Indeed, in that case,

for  $4\lambda\nu^3 > 1$  the rate  $r(\nu)$  is  $(2\nu)^{-1}$  and thus is decreasing, while for  $4\lambda\nu^3 < 1$ ,

$$\begin{aligned} r(\nu) &= \frac{1}{2\nu} \left( 1 - \sqrt{1 - 4\lambda\nu^3} \right) \\ \Rightarrow r'(\nu) &= \frac{r(\nu)}{\nu} \left( -1 + \frac{6\lambda\nu^3}{\sqrt{1 - 4\lambda\nu^3} (1 - \sqrt{1 - 4\lambda\nu^3})} \right) \\ &= \frac{r(\nu)}{\nu} \left( \frac{1}{2} + \frac{3}{2\sqrt{1 - 4\lambda\nu^3}} \right) \end{aligned}$$

and thus is increasing. Hence, for a homothety  $S = \lambda I_N$ , the rate of convergence to equilibrium in (6) is maximal for  $\nu = (4\lambda)^{-\frac{1}{3}}$ , and in that case this optimal rate is  $(\lambda/2)^{\frac{1}{3}}$ . In comparison, the rate of convergence of

$$dX_t = -\lambda X_t dt + \sqrt{2} dB_t$$

is  $\lambda$ , which is better than  $(\lambda/2)^{\frac{1}{3}}$  if and only if  $\lambda > \frac{1}{\sqrt{2}} \simeq 0.7$ .

### 3 Proofs

Let us start with an easy lemma which will enable us to characterize  $A$  in the couple  $(A, D) \in \mathcal{I}(S)$  once  $D$  is fixed.

**Lemma 3.** *For  $S \in \mathcal{S}_N^{>0}(\mathbb{R})$ , the following are equivalent:*

- $(A, D) \in \mathcal{I}(S)$
- $D \in \mathcal{S}_N^{\geq 0}(\mathbb{R})$  with  $\text{Tr} D \leq N$  and there exists  $J \in \mathcal{A}_N(\mathbb{R})$  such that

$$A = -(D + J)S.$$

*Proof.* Fix  $D \in \mathcal{S}_N^{\geq 0}(\mathbb{R})$  such that  $\text{Tr} D \leq N$ . First we see that

$$L'_{-DS,D}(\psi_\infty) = \nabla \cdot ((DSx)\psi_\infty - (DSx)\psi_\infty) = 0.$$

As a consequence

$$(L'_{A,D} - L'_{-DS,D})\psi_\infty = -\nabla \cdot ((A + DS)x\psi_\infty)$$

and [10, Lemma 1] concludes.  $\square$

We may now proceed to the proof of our main result.

*Proof of Theorem 1.* Let  $(A, D) \in \mathcal{I}(S)$ , so that by the previous lemma 3,  $D \in \mathcal{S}_N^{\geq 0}(\mathbb{R})$  such that  $\text{Tr} D \leq N$ ,  $J \in \mathcal{A}_N(\mathbb{R})$  and  $A = -(D + J)S$ . Following [10] we write

$$(D + J)S = (S)^{-\frac{1}{2}} (\tilde{D} + \tilde{J}) (S)^{\frac{1}{2}}$$

where  $\tilde{D} = (S)^{\frac{1}{2}} D (S)^{\frac{1}{2}}$  and similarly for  $\tilde{J}$ . Note that  $M \leftrightarrow (S)^{\frac{1}{2}} M (S)^{\frac{1}{2}}$  is a bijection which leaves  $\mathcal{S}_N^{\geq 0}(\mathbb{R})$  and  $\mathcal{A}_N(\mathbb{R})$  invariant, and that

$$\rho(A) = \rho(-\tilde{D} - \tilde{J}).$$

From [10, Propositions 1 and 4],

$$\max \left\{ \rho(-\tilde{D} - \tilde{J}), \tilde{J} \in \mathcal{A}_N(\mathbb{R}) \right\} = \frac{\text{Tr} \tilde{D}}{N}.$$

One may object that [10, Propositions 1 and 4] are written for an invertible matrix, which is not necessarily the case for  $\tilde{D}$ . It can be seen that this restriction is not necessary in the proof; or to save the reader from a careful check of these proofs, we may note that  $\tilde{D} + \varepsilon I_N$  falls within the scope of [10], which yields the same result. Hence

$$\max \{ \rho(A), (A, D) \in \mathcal{I}(S) \} = \frac{1}{N} \max \left\{ \text{Tr} \left( (S)^{\frac{1}{2}} D (S)^{\frac{1}{2}} \right), D \in \mathcal{S}_N^{\geq 0}(\mathbb{R}), \text{Tr} D \leq N \right\}.$$

Let  $Q$  be an orthonormal matrix such that  $S^{\frac{1}{2}} = Q^T \Sigma Q$  with  $\Sigma = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_N})$ .

$$\begin{aligned} \text{Tr} \left( (S)^{\frac{1}{2}} D (S)^{\frac{1}{2}} \right) &= \text{Tr} (\Sigma Q D Q^T \Sigma) \\ &\leq \max \{ \lambda_i, i = 1..N \} \text{Tr} (Q D Q^T) \\ &= \max \sigma(S) \text{Tr} (D). \end{aligned}$$

We have proved

$$\max \{ \rho(A), (A, D) \in \mathcal{I}(S) \} \leq \max \sigma(S)$$

and to prove equality holds we need to find  $D \in \mathcal{S}_N^{\geq 0}(\mathbb{R})$  with  $\text{Tr} D \leq N$  such that  $\text{Tr} \tilde{D} = N \max \sigma(S)$ . Let  $i \in \llbracket 1, N \rrbracket$  be such that  $\lambda_i = \max \sigma(S)$ , let  $E_{i,i}$  be the  $N \times N$  matrix with all coefficients being zero except the coefficient  $(i, i)$  being equal to 1, and set  $D = N Q^T E_{i,i} Q$ . In other words  $D = N v_i v_i^T$  where  $v_i$  is a normalized eigenvector of  $S$  associated to  $\lambda_i$ . In that case

$$\begin{aligned} \tilde{D} &= N Q^T \Sigma Q Q^T E_{i,i} Q Q^T \Sigma Q \\ &= \lambda_i D \end{aligned}$$

which concludes the proof.  $\square$

*Proof of Theorem 2.* Without loss of generality (up to an orthonormal change of variables) we can assume the vector  $v$  with all coordinates being equal to 1 is an eigenvector of  $S$  associated to  $\lambda = \max \sigma(S)$ . Let  $D = v v^T$ , which is the matrix with all coefficients equal to 1. Let  $0 < \nu_1 < \nu_2 < \dots < \nu_N$  (to be specified later) and  $Q = \text{diag}(\nu_1, \dots, \nu_N)$ . Note that the eigenvectors of  $Q$  are obviously the canonical basis vectors  $(e_1, \dots, e_N)$ , which satisfy  $e_i^T D e_i = 1 = \text{Tr} D / N$  for all  $i$ . Let  $\tilde{J}$  be the antisymmetric matrix defined by

$$(\tilde{J})_{k,l} = \frac{\nu_k + \nu_l}{\nu_k - \nu_l}$$

if  $k \neq l$  and 0 else. According to [10, Lemma 2 and Equation (38)],

$$Q \tilde{J} - \tilde{J} Q = -D Q - Q D + 2Q \tag{7}$$

and  $\sigma(D + \tilde{J}) \subset 1 + i\mathbb{R}$ . Let

$$A = -\lambda S^{-\frac{1}{2}} (D + \tilde{J}) S^{\frac{1}{2}} = -\left( D + S^{-\frac{1}{2}} \tilde{J} S^{-\frac{1}{2}} \right) S,$$

so that  $(A, D) \in \mathcal{I}(S)$  according to Lemma 3. Recall that  $L_{A,D}^* = L_{C,D}$  with

$$\begin{aligned} C &:= -2DS - A \\ &= -\lambda S^{-\frac{1}{2}} (D - \tilde{J}) S^{\frac{1}{2}}. \end{aligned}$$

and note that  $\sigma(D - \tilde{J}) = \sigma(D + \tilde{J})^T \subset 1 + i\mathbb{R}$ .



Let  $\alpha$  be the function  $r > 0 \mapsto \alpha(r) = r \ln r$  (or  $r \in \mathbb{R} \mapsto \alpha(r) = \frac{1}{2}r^2$  if one desires to deal with  $L^2$  decay rather than entropy), so that  $\alpha''(r)$  is  $r^{-1}$  (or 1). According to [12, Lemma 8], for all  $M \in \mathcal{S}_N^{>0}(\mathbb{R})$  and for all  $h > 0$ , denoting by  $h_t = e^{tL_{C,D}}h$ ,

$$\partial_t \left( \alpha''(h_t) (\nabla h_t)^T M \nabla h_t \right) \leq 2\alpha''(h_t) (\nabla h_t)^T M C^T \nabla h_t$$

Applying this with  $M = S^{-\frac{1}{2}}Q S^{-\frac{1}{2}}$ , we obtain

$$\begin{aligned} \partial_t \left( \alpha''(h_t) \left| Q^{\frac{1}{2}} S^{-\frac{1}{2}} \nabla h_t \right|^2 \right) &\leq -2\lambda \alpha''(h_t) (\nabla h_t)^T S^{-\frac{1}{2}} Q (D + \tilde{J}) S^{-\frac{1}{2}} \nabla h_t \\ &= -\lambda \alpha''(h_t) \left( S^{-\frac{1}{2}} \nabla h_t \right)^T \left( QD + DQ + Q\tilde{J} - \tilde{J}Q \right) S^{-\frac{1}{2}} \nabla h_t \\ &= -2\lambda \alpha''(h_t) \left( S^{-\frac{1}{2}} \nabla h_t \right)^T Q S^{-\frac{1}{2}} \nabla h_t \end{aligned}$$

where the first equality comes from the fact that  $Q$  and  $D$  are symmetric and  $\tilde{J}$  is antisymmetric, and the last equality from (7). Hence

$$\alpha''(h_t) \left| Q^{\frac{1}{2}} S^{-\frac{1}{2}} h_t \right|^2 \leq e^{-2\lambda(t-s)} \alpha''(h_s) \left| Q^{\frac{1}{2}} S^{-\frac{1}{2}} h_s \right|^2.$$

The log-Sobolev inequality for the standard Gaussian distribution  $\gamma(dx) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}|x|^2} dx$  reads

$$\text{Ent}_\gamma f \leq \frac{1}{2} \int \frac{|\nabla f|^2}{f} d\gamma,$$

for all  $f > 0$  such that the r.h.s is finite. By the change of variable  $z = S^{\frac{1}{2}}x$  it yields

$$\text{Ent}_{\psi_\infty} f \leq \frac{1}{2} \int \frac{(\nabla f)^T S^{-1} \nabla f}{f} d\psi_\infty.$$

Now

$$\begin{aligned} \text{Ent}_{\psi_\infty}(h_t) &\leq \frac{1}{2} \int \frac{(\nabla h_t)^T S^{-1} \nabla h_t}{h_t} d\psi_\infty \\ &\leq \frac{1}{2\nu_1} \int \frac{\left| Q^{\frac{1}{2}} S^{-\frac{1}{2}} \nabla h_t \right|^2}{h_t} d\psi_\infty \\ &\leq \frac{e^{-2\lambda(t-s)} \nu_N}{2\nu_1} \int \frac{\left| S^{-\frac{1}{2}} \nabla h_s \right|^2}{h_s} d\psi_\infty, \\ &\leq \frac{\nu_N}{2\nu_1 \min \sigma(S)} e^{-2\lambda(t-s)} \int \frac{|\nabla h_s|^2}{h_s} d\psi_\infty. \end{aligned}$$

Finally the elliptic reversible generator  $L_{-S, I_N}$  satisfies the Bakry-Emery criterion  $\Gamma_2 \geq 0$  (see [3] for instance, more precisely [3, Equation 1.16.5 p. 72] together with [3, Theorem 5.5.2.(v), p.259] with  $\rho = 0$  and integrated with respect to  $\psi_\infty$ ), so that if now  $h_t = e^{tL_{-S, I_N}}h$ ,

$$\int \frac{|\nabla h_s|^2}{h_s} d\psi_\infty \leq \frac{1}{s} \text{Ent}_{\psi_\infty} h.$$

We can then take  $\nu_N$  arbitrarily close to  $\nu_1$  to get the first part of Theorem 2. On the other hand, following [10, Remark 8], if we choose  $\nu_k = N + k$  (so that  $\nu_N \leq 2\nu_1$ ), using that for any

$T \in \mathcal{S}_N^{\geq 0}(\mathbb{R})$ ,  $\|AT\|_F \leq \max \sigma(T)\|A\|_F$ , we get

$$\begin{aligned}\|A\|_F^2 &= \lambda^2 \|S^{-\frac{1}{2}} (D + \tilde{J}) S^{\frac{1}{2}}\|_F^2 \\ &\leq \frac{\lambda^3}{\min \sigma(S)} \left( N + \sum_{j \neq k} \left( \frac{\nu_k + \nu_j}{\nu_k - \nu_j} + 1 \right)^2 \right) \\ &\leq \frac{\lambda^3}{\min \sigma(S)} (N + N(N-1)(4N)^2) \\ &\leq \frac{16\lambda^3 N^4}{\min \sigma(S)}.\end{aligned}$$

□

Note that an optimal  $(A, D)$  is explicitly constructed in this proof. The procedure may be decomposed in two steps: first, exhibit an eigenvector  $v$  of  $S$  associated to  $\max \sigma(S)$ , and set  $D = vv^T$ . The second step answers the following general question: given any  $D \in \mathcal{S}_N^{\geq 0}(\mathbb{R})$ , and writing  $\tilde{D} = (S)^{\frac{1}{2}} D (S)^{\frac{1}{2}}$ , how to construct an optimal  $\tilde{J} \in \mathcal{A}_N(\mathbb{R})$  so that

$$\rho(-\tilde{D} - \tilde{J}) = \frac{\text{Tr} \tilde{D}}{N} ?$$

By a straightforward adaptation of [10, Algorithm p. 252] (just change  $S$  by  $\tilde{D}$  everywhere), construct an orthonormal basis  $(\psi_1, \dots, \psi_N)$  such that  $\psi_i^T \tilde{D} \psi_i = \text{Tr} \tilde{D} / N$  for all  $i$ . Let  $P$  be the matrix whose columns are the  $\psi_i$ 's,  $\nu_k = N + k$  for  $k \in \llbracket 1, N \rrbracket$  and  $\hat{J}$  be the antisymmetric matrix with coefficients  $\frac{\nu_k + \nu_l}{\nu_k - \nu_l}$  for  $k \neq l$ . Then

$$\tilde{J} := P \hat{J} P^{-1}$$

is a solution of the problem, and we conclude by setting  $A = -DS - (S)^{-\frac{1}{2}} \tilde{J} (S)^{\frac{1}{2}}$ .

## 4 Numerical illustrations

In dimension 2, consider  $S = \text{diag}(\varepsilon, 1)$ . For any  $h \in \mathbb{R}$ , the corresponding  $\psi_\infty$  is the unique equilibrium of

$$dX_t = - \begin{pmatrix} \varepsilon & -h \\ \varepsilon h & 1 \end{pmatrix} X_t dt + \sqrt{2} dB_t \quad (8)$$

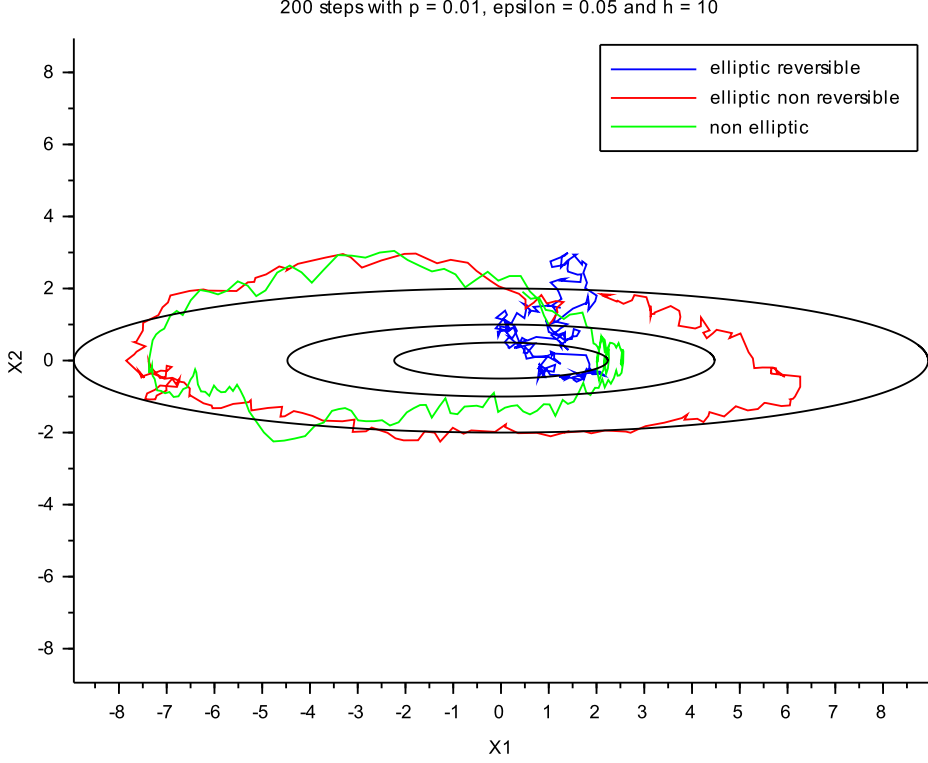
where  $B = (B^1, B^2)$  is a 2-dimensional Brownian motion. It is also an equilibrium for

$$dZ_t = - \begin{pmatrix} 0 & -h \\ \varepsilon h & 2 \end{pmatrix} Z_t dt + \sqrt{2} \begin{pmatrix} 0 \\ dB^1 + dB^2 \end{pmatrix} \quad (9)$$

which is hypoelliptic as soon as  $h \neq 0$ . When  $h$  is large enough, away from the origin (so that the random forces are small with respect to the deterministic drift), the behaviours of  $X$  and  $Z$  are similar, mostly driven by a fast rotation, while the first coordinate of the reversible process solving (8) with  $h = 0$  moves slower, and thus covers the space less efficiently (see Fig. 1, where the parameter  $p$  is the step size of the Euler Scheme). Note that in the case of  $Z$ , even if this rotation is randomly perturbed, since  $dZ^1 = hZ^2 dt$ , the process always goes from left to right in the lower half-plane  $\{(x, y), y < 0\}$  and from right to left in the upper one.

In (8), the optimal rate  $\frac{1+\varepsilon}{2}$  is obtained for  $h^2 \geq \frac{(1+\varepsilon)^2}{4\varepsilon} - 1$  while in (9) the optimal rate 1 is obtained for  $h^2 \geq \frac{1}{\varepsilon}$ . For instance if we chose  $h = \sqrt{\frac{2}{\varepsilon}}$ , then both conditions are fulfilled and in both cases the drift matrix is diagonalizable with two conjugated distinct eigenvalues. For

Figure 1: Trajectories for a multi-scale equilibrium.



a diagonalizable  $2 \times 2$  matrix  $A$  with eigenvalues  $\lambda_1 = \bar{\lambda}_2 \neq \lambda_2$ , denoting by  $\nu = |\lambda_1 - \lambda_2|$  and by  $\alpha = |\bar{v}_1^T v_2|^{-1}$  where  $(v_1, v_2)$  is a normalized eigenbasis of  $A$ , the Hermitian matrix norm of  $e^{tA}$  can be explicitly computed (see e.g. [14, Lemma 3]) as

$$\|e^{tA}\|^2 = e^{2\Re(\lambda_1)t} \left( 1 + \frac{2}{\sqrt{\frac{2(\alpha^2-1)}{1-\cos(\nu t)} + 1} - 1} \right).$$

This is represented in Figure 2 (with  $\epsilon = 0.05$  for every curves,  $h = \sqrt{\frac{2}{\epsilon}}$  for the second and third ones and  $h = \sqrt{\frac{1}{\epsilon}}$  for the last one). A large  $h$  seems to improve the prefactor, and indeed, note that

$$M(h) := \max_{t \geq 0} \left( 1 + \frac{2}{\sqrt{\frac{2(\alpha^2-1)}{1-\cos(\nu t)} + 1} - 1} \right) = 1 + \frac{2}{\alpha - 1}$$

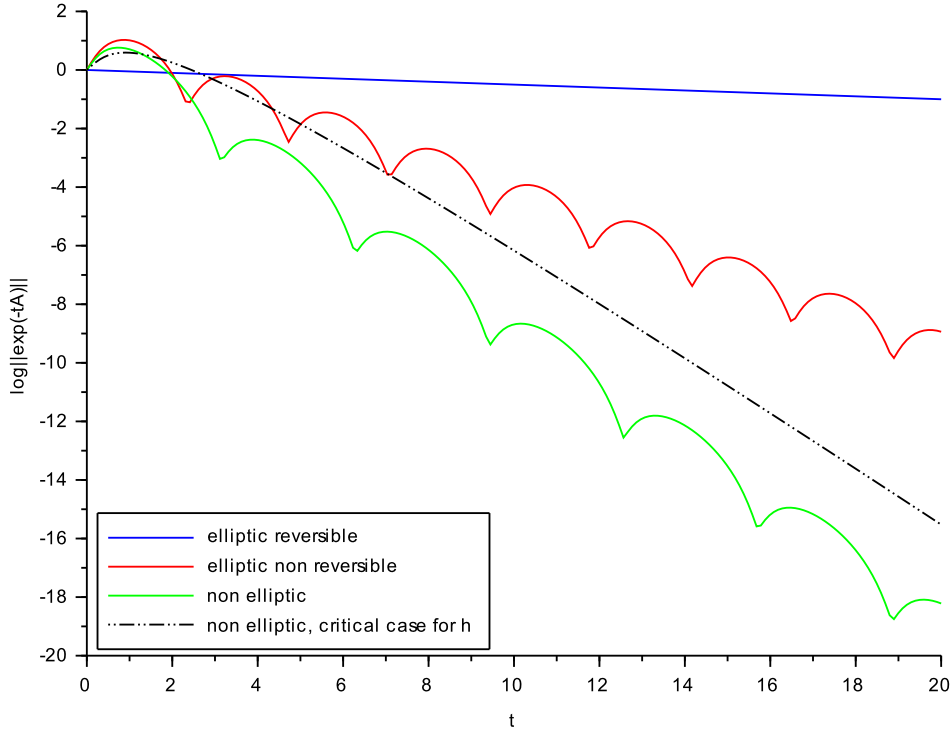
and that, normalizing the eigenvectors  $v_j = (-h, \lambda_j)$  for  $j = 1, 2$ , we can compute

$$\alpha^{-2} = \frac{\left| h^2 + \left( 1 + i\sqrt{h^2\epsilon - 1} \right)^2 \right|^2}{(h^2 + 1 + h^2\epsilon - 1)^2} = \frac{(1 - \epsilon)^2 + 4h^{-2}}{(1 + \epsilon)^2},$$

which decreases with  $h^2$ , together with the prefactor. As  $h$  goes to infinity,  $M(h)$  goes to its minimum, which is  $\frac{1}{\epsilon}$  (if  $\epsilon \leq 1$ ).

As remarked in the Introduction, how to boost the speed of convergence of a Markov process to sample a Gibbs measure is perhaps less relevant for quadratic potentials. There

Figure 2: Norms of the drift matrix exponentials



are thus two distinct problems that we have in mind for the future. The first one deals with the direct generalization of our result when we replace the Gaussian measure  $\psi_\infty$  by  $e^{-V}$  where  $V$  satisfies  $Hess(V) \geq S > 0$ , with  $S$  constant positive symmetric. Is it possible to add a divergence free drift and a potentially degenerate (constant) diffusion matrix so that the rate of convergence to equilibrium is  $\max(\sigma(S))$ ? The question is also of great interest when the dynamics is metastable, namely when  $V$  has several local minima. A toy problem of this phenomenon would be to consider in dimension 1 (even if MCMC algorithm usually outperforms deterministic algorithms only in large dimension) a two-wells potential

$$V(x) = ax^4 - bx^2$$

with  $a, b > 0$ . Then  $V$  has two minima  $-b^2/(4a)$  attained at  $x = \pm\sqrt{b/(2a)}$  and separated by a local maxima 0 at  $x = 0$ . Depending on the energy barrier  $V(0) - V\left(\sqrt{b/(2a)}\right) = b^2/(4a)$  to overcome in order to go from one catchment area to the other, the reversible Fokker-Planck diffusion (1) will take a long time to achieve such a crossing. In Figures 3 and 4 are represented two such trajectories over different periods, along with trajectories of the first coordinate of a kinetic Langevin diffusion (3) (in each case both the reversible and the kinetic diffusions are driven by the same Brownian motion). As discussed at the end of Section 2, for the Langevin process there should be another parameter to tune, the variance  $\nu$  of the velocity at equilibrium. Since the optimal choice of  $\nu$  is already non trivial in the quadratic case, we won't address it here in this metastable context: for now, our considerations are only qualitative.

The first point to comment in Figure 3 is that the trajectory is smoother in the kinetic case than in the reversible one, which is obvious since in the first case it is 3/2-Hölder continuous while in the second one it is only 1/2-Hölder continuous. Second, due to its inertia, the trajectory in the kinetic case shows large oscillations in which kinetic and potential energy successively convert one to the other. In particular from the times  $t \simeq 6.5$  to 10 we can see the

Figure 3: Metastable trajectories in short time.

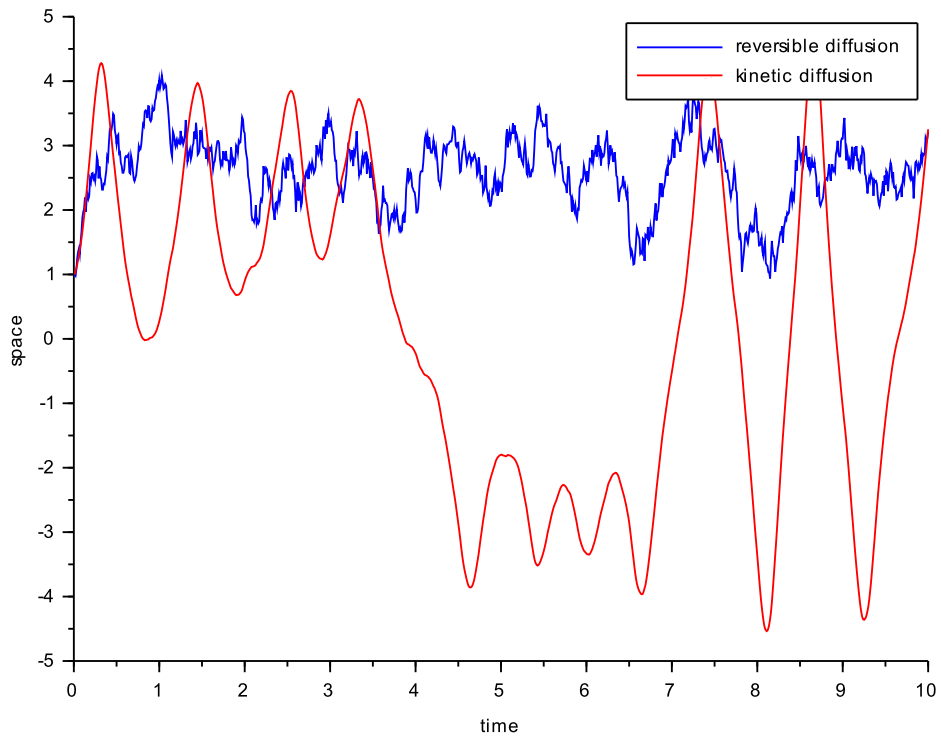
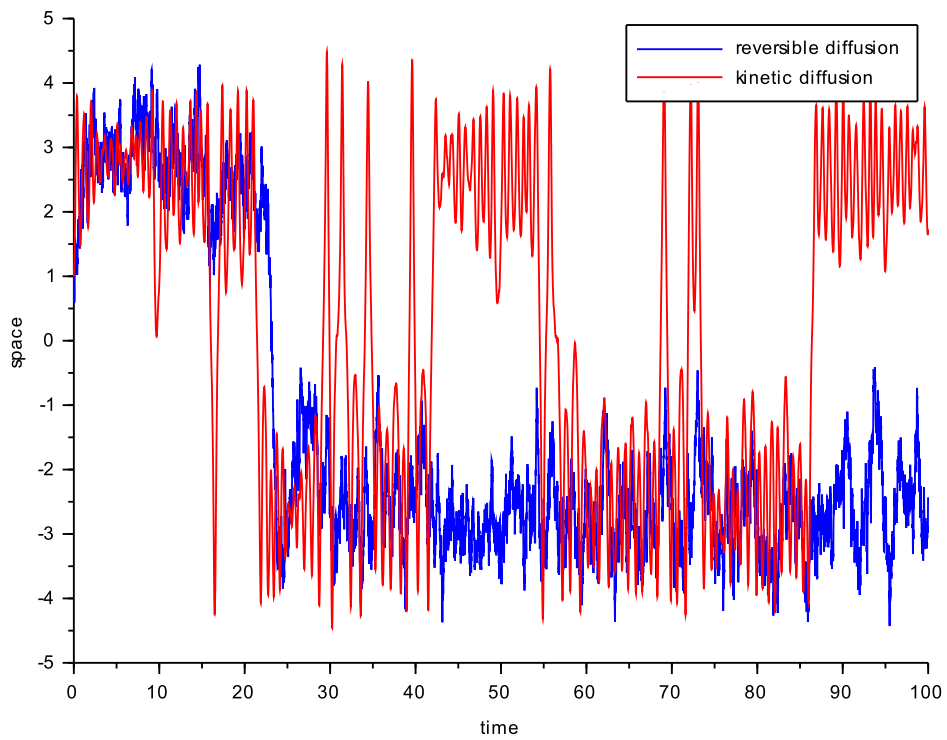


Figure 4: Metastable trajectories in longer time.



process has a high level of total energy and thus these large oscillations cross the energy barrier at  $x = 0$  without difficulty. At some point the total energy will decrease sufficiently for the process to stay trapped in the vicinity of one of the two minima, which has then a reasonable chance to be different from the one from which it started before the energy level got high.

That way we would interpret Figure 4 as an illustration to the fact the Langevin dynamics deals more efficiently with metastability (or at least energy barriers) than the reversible Fokker-Planck diffusion. With Theorem 2 in mind, we could also interpolate from these figures the behaviour of a process that switch at random times from Equation (8) to (9) (or anything else in that spirit). However it is difficult to export an intuition based on a toy model in dimension 1 and with a fixed set of parameter (especially the variance of the velocity in (9)) to a more general case.

### Acknowledgements

The authors would like to thank the referees for well pointed remarks which have led to a significant improvement of the presentation of the paper.

## References

- [1] A. Arnold, E. Carlen, and Q. Ju. Large-time behavior of non-symmetric Fokker-Planck type equations. *Commun. Stoch. Anal.*, 2(1):153–175, 2008.
- [2] A. Arnold and J. Erb. Sharp entropy decay for hypocoercive and non-symmetric Fokker-Planck equations with linear drift. *ArXiv e-prints*, September 2014.
- [3] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators*, volume 348 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2014.
- [4] F. Bolley and I. Gentil. Phi-entropy inequalities for diffusion semigroups. *J. Math. Pures Appl. (9)*, 93(5):449–473, 2010.
- [5] P. Diaconis and L. Miclo. On the spectral analysis of second-order Markov chains. *Ann. Fac. Sci. Toulouse Math. (6)*, 22(3):573–621, 2013.
- [6] S. Gadat and L. Miclo. Spectral decompositions and  $\mathbb{L}^2$ -operator norms of toy hypocoercive semi-groups. *Kinet. Relat. Models*, 6(2):317–372, 2013.
- [7] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte-Carlo methods. *J. Royal. Stat. Soc., Series B*, 73(2):1–37, 2011.
- [8] C.R. Hwang, S.Y. Hwang-Ma and S.J. Sheu. Accelerating Gaussian diffusions. *Ann. Appl. Probab.*, 3(3):897–913, 1993.
- [9] C.R. Hwang, S.Y. Hwang-Ma and S.J. Sheu. Accelerating diffusions. *Ann. Appl. Probab.*, 15(2):1433–1444, 2005.
- [10] T. Lelièvre, F. Nier, and G. A. Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *Journal of Statistical Physics*, 152(2):237–274, 2013.
- [11] T. Lelièvre, M. Rousset, and G. Stoltz. *Free energy computations: A mathematical perspective*. Imperial College Press, 2010.
- [12] P. Monmarché. Generalized  $\Gamma$  calculus and application to interacting particles on a graph. *ArXiv e-prints*, October 2015.
- [13] P. Monmarché. Hypocoercivity in metastable settings and kinetic simulated annealing. *ArXiv e-prints*, February 2015.
- [14] L. Miclo, P. Monmarché. Étude spectrale minutieuse de processus moins indécis que les autres. *Lecture Notes in Mathematics*, September 2012.

- [15] A. Scemama, T. Lelièvre, G. Stoltz, and M. Caffarel. An efficient sampling algorithm for variational monte carlo. *Journal of Chemical Physics*, 125, September 2006.